

# Semiparametric modeling and estimation of heteroscedasticity in regression analysis of cross-sectional data

Ingrid Van Keilegom

*Institute of Statistics, Université catholique de Louvain, Voie du Roman Pays 20,  
1348 Louvain-la-Neuve, Belgium  
e-mail: [ingrid.vankeilegom@uclouvain.be](mailto:ingrid.vankeilegom@uclouvain.be)*

and

Lan Wang

*School of Statistics, University of Minnesota, 224 Church Street SE,  
Minneapolis, MN 55455 USA  
e-mail: [lan@stat.umn.edu](mailto:lan@stat.umn.edu)*

**Abstract:** We consider the problem of modeling heteroscedasticity in semiparametric regression analysis of cross-sectional data. Existing work in this setting is rather limited and mostly adopts a fully nonparametric variance structure. This approach is hampered by *curse of dimensionality* in practical applications. Moreover, the corresponding asymptotic theory is largely restricted to estimators that minimize certain smooth objective functions. The asymptotic derivation thus excludes semiparametric quantile regression models. To overcome these drawbacks, we study a general class of location-dispersion regression models, in which both the location function and the dispersion function are semiparametrically modeled. We establish unified asymptotic theory which is valid for many commonly used semiparametric structures such as the partially linear structure and single-index structure. We provide easy to check sufficient conditions and illustrate them through examples. Our theory permits non-smooth location or dispersion functions, thus allows for semiparametric quantile heteroscedastic regression and robust estimation in semiparametric mean regression. Simulation studies indicate significant efficiency gain in estimating the parametric component of the location function. The results are applied to analyzing a data set on gasoline consumption.

**Keywords and phrases:** Dispersion function, heteroscedasticity, partially linear model, quantile regression, semiparametric regression, single-index model, variance function.

Received November 2009.

## 1. Introduction

The problem of heteroscedasticity, which traditionally means nonconstant variance function, frequently arises in regression analysis of economic data. In this

paper, we broaden the scope of heteroscedasticity by considering a general class of location-dispersion regression models, where the relation between a response variable  $Y$  and a covariate vector  $X$  is given by

$$Y = m(X) + \sigma(X)\varepsilon. \quad (1.1)$$

In the above model,  $\varepsilon$  denotes the random error,  $m(\cdot)$  is called *regression function* and the nonnegative function  $\sigma(\cdot)$  is called *dispersion function*. With different specifications on  $\varepsilon$ , this formulation includes both the conditional mean and conditional median (or more general quantile) regression models.

In the literature, model (1.1) has been thoroughly investigated for parametric mean regression, where  $m(\cdot)$  is characterized by a finite-dimensional parameter and  $E(\varepsilon|X) = 0$  (Davidian, Carroll and Smith, 1988; Zhao, 2001; Chapter 11, Greene, 2002, among others). This paper focuses on semiparametric regression models, which are less studied but are extremely useful due to their flexibility to accommodate nonlinearity and to circumvent curse of dimensionality (Härdle, Liang and Gao, 2000; Ruppert, Wand and Carroll, 2003; Yatchew, 2003). In particular, we consider the general setup with  $m(x) = m(x, \alpha_0, r_0)$ , where  $\alpha_0$  is a finite dimensional parameter and  $r_0$  is an infinite dimensional parameter. The main interest is often in making inference about  $\alpha_0$  while treating  $r_0$  as a nuisance parameter, which can only be estimated at a slower than  $\sqrt{n}$  nonparametric rate.

In semiparametric regression models, the commonly used estimation procedures in general still yield consistent estimators for  $\alpha_0$  even if heteroscedasticity is not accounted for. However, efficiency loss due to ignoring heteroscedasticity may be substantial. Moreover, the correctness of the standard error formula for  $\alpha_0$  and the validity of the associated confidence intervals or hypothesis testing procedures depend critically on the dispersion function (Akritas and Van Keilegom, 2001; Carroll, 2003). In addition, it is sometimes important to model the dispersion function in order to obtain a satisfactory bandwidth for estimating the nonparametric part of the regression function. Ruppert et al. (1997) provided such an example, where the heteroscedasticity is severe and the variance function has to be estimated in order to obtain a good bandwidth for estimating the derivative of the mean function.

In the semiparametric regression setting, Schick (1996), Liang, Härdle and Carroll (1999), Härdle, Liang and Gao (2000, §2), Ma, Chiou and Wang (2006) have studied heteroscedastic partially linear mean regression models, where the variance function  $\sigma^2(x)$  is assumed to be smooth but unknown. They estimate the variance function nonparametrically, and then use the estimator to construct weights to achieve more efficient estimation of the parametric component of the mean regression function. Härdle, Hall and Ichimura (1993) investigated heteroscedastic single-index models, so did Xia, Tong and Li (2002); and Chiou and Müller (2004) proposed a flexible semiparametric quasi-likelihood, which assumes that the mean function has a multiple-index structure and the variance function has an unknown nonparametric form.

The aforementioned work, however, suffers from several drawbacks. First, they have all adopted a fully nonparametric model for the variance function.

This approach does not work well in high dimension due to curse of dimensionality. Second, their asymptotic theory can only be applied to a specific semiparametric structure. Third, these methods require a smooth objective function thus do not apply to semiparametric quantile regression models, see for instance [He and Liang \(2000\)](#), [Lee \(2003\)](#), [Horowitz and Lee \(2005\)](#). In fact, existing study of heteroscedastic quantile regression is restricted to parametrically specified quantile function ([Koenker and Zhao, 1994](#); [Zhao, 2001](#)). The last point is also relevant when one is interested in robust estimation for the mean regression model in the presence of outlier contamination.

We also like to mention [Müller and Zhao \(1995\)](#), who consider a general semiparametric variance function model in a fixed design regression setting. In their model, the regression function is assumed to be smooth and is modeled nonparametrically, whereas the relation between the variance and the mean regression function is assumed to follow a generalized linear model. However, although the variance function has both a parametric and nonparametric component, and so can be considered as being semiparametric, its model differs quite a bit from the semiparametric model we use in this paper.

The above concerns motivate us to propose a flexible semiparametric framework for modeling heteroscedasticity and to develop a unified theory that applies to general semiparametric structures and non-smooth objective functions. In particular, we advocate to adopt a semiparametric structure for modeling the dispersion function. This approach avoids the rigid assumption imposed by a parametric dispersion function; at the same time it circumvents the curse of dimensionality introduced by a nonparametric dispersion function. In this general framework, we establish an asymptotic normality theory for estimating the form of heteroscedasticity by building on the work of [Chen, Linton and Van Keilegom \(2003\)](#), who developed a general theory for semiparametric estimation with a non-smooth criterion function. We provide a set of easy to check sufficient conditions, such that the asymptotic normality theory is valid for many commonly used semiparametric structures, for instance, the partially linear structure and the single-index structure. We discuss but do not get deep into how the knowledge of heteroscedasticity can be used to construct a more efficient weighted estimator for the parametric component of  $m(\cdot)$ .

We discuss two different constraints for the random error  $\varepsilon$  in (1.1): the mean zero constraint and the median zero constraint, which correspond to mean regression and median regression, respectively. Although the current theory is restricted to cross-sectional data, the ideas and techniques can be applied to extend to time-series models for heteroscedastic economic and financial data, such as the autoregressive conditional heteroscedastic (ARCH) model of [Engle \(1982\)](#).

The paper is organized as follows. In Section 2 we formally introduce the semiparametric location-dispersion model and discuss how to estimate the dispersion function. Section 3 provides generic assumptions that are applicable to general semiparametric models, and presents the asymptotic normality theory for estimating the dispersion function. In Section 4 we verify these generic conditions for two particular semiparametric models. The finite sample behavior of

the proposed methods is examined in Section 5, while Section 6 is devoted to the analysis of data on gasoline consumption. In Section 7 some ideas for future research are discussed. Finally, all proofs are collected in the Appendix.

## 2. Estimation of a semiparametric dispersion function

### 2.1. Semiparametric location-dispersion model

We consider a general semiparametric location-dispersion model:

$$Y = m(X, \alpha_0, r_0) + \sigma(X, \beta_0, g_0)\varepsilon, \quad (2.1)$$

where  $X = (X_1, \dots, X_d)^T$  is a  $d$ -dimensional covariate vector with compact support  $R_X$ ,  $\alpha_0$  and  $\beta_0$  are finite dimensional parameters, and  $r_0$  and  $g_0$  are infinite dimensional parameters. Let  $(X_i^T, Y_i)^T = (X_{1i}, \dots, X_{di}, Y_i)^T$  be i.i.d. copies of  $(X^T, Y)^T$ . The conditions that need to be imposed on  $\varepsilon$  to make the model identifiable are given in Section 2.3 (mean regression) and Section 2.4 (median regression).

The dispersion function is assumed to have a general semiparametric structure. This paper discusses two examples in detail (Section 4), corresponding to the exponentially transformed partially linear structure  $\sigma(X, \beta_0, g_0) = \exp(\beta_0^T X_{(1)} + g_0(X_{(2)}))$  with  $X = (X_{(1)}^T, X_{(2)}^T)^T$  and the single-index structure  $\sigma(X, \beta_0, g_0) = g_0(\beta_0^T X)$ , respectively. We assume that the unknown function  $g_0$  belongs to some space  $\mathcal{G}$  of uniformly bounded functions that depend on  $X$  and  $\beta$  through a variable  $U = U(X, \beta)$ , where  $\beta$  belongs to a compact set  $B$  in  $\mathbb{R}^\ell$ , with  $\ell \geq 1$  depending on the model (e.g.  $U(X, \beta) = X_{(2)}$  and  $U(X, \beta) = \beta^T X$  for the above partial linear and single index structures respectively). For any function  $g$ , the notation  $g_\beta$  will be used to indicate the (possible) dependence on  $\beta$ . The estimator of the true  $g_0$  will in fact in many situations be a profile estimator, depending on  $\beta$  (see the examples in Section 4). For notational convenience we use the abbreviated notation  $(\beta, g) = (\beta, g_\beta(\cdot))$ ,  $(\beta, g_0) = (\beta, g_{0\beta}(\cdot))$  and  $(\beta_0, g_0) = (\beta_0, g_{0\beta_0}(\cdot))$ , whenever no confusion is possible. Whenever needed, we will replace  $\sigma(X, \beta, g)$  by  $\sigma(X, \beta, g_\beta)$  or  $\sigma(X, \beta, g_\beta(U))$  to highlight the dependence of the function  $g$  on the parameter  $\beta$  or on the variable  $U$  (note that this implies that the third argument of the function  $\sigma$  can be a function in  $\mathcal{G}$  or an element of  $\mathbb{R}$ , depending on which notation we use).

To keep the notations and presentation simple, we assume that both  $g_0$  and  $U$  are one-dimensional. However, all the results in this paper can be extended in a straightforward way to the multi-dimensional case. For example, we may have  $g_0 = (g_{01}, \dots, g_{0k})$  for some  $k \geq 1$ , which allows a multiplicative model for  $\sigma$  of the form  $\sigma(x) = \prod_{j=1}^d g_{0j}(x_j)$ .

Although we will discuss later how to use the estimated dispersion function to construct a more efficient estimator for  $\alpha_0$ , our main interest in this paper is to establish a general theory for estimating  $\beta_0$  and  $\sigma(X, \beta_0, g_0)$ . Therefore, we will simply write  $m_0$  or  $m_0(X)$  to denote the regression function in the sequel.

## 2.2. A motivating example

To help understand the general estimation procedure, we start with a motivating example given by

$$Y = m_0(X) + \exp\{\beta_0^T X_{(1)} + g_0(X_{(2)})\}\varepsilon, \quad (2.2)$$

where  $E(\varepsilon|X) = 0$ ,  $\text{Var}(\varepsilon|X) = 1$  and  $X = (X_{(1)}^T, X_{(2)}^T)^T$ , with  $X_{(1)} = (X_1, \dots, X_{d-1})^T$  and  $X_{(2)} = X_d$ . Note that model (2.2) can also be written as

$$\begin{aligned} (Y - m_0(X))^2 &= \exp\left\{2\beta_0^T X_{(1)} + 2g_0(X_{(2)})\right\} \\ &\quad + \exp\left\{2\beta_0^T X_{(1)} + 2g_0(X_{(2)})\right\}(\varepsilon^2 - 1). \end{aligned} \quad (2.3)$$

Since  $E(\varepsilon^2 - 1|X) = 0$ , the regression function in the new model (2.3) is equal to the variance function in the original model (2.2). Therefore we can apply estimation techniques known from the literature on semiparametric regression estimation to estimate the variance function. In particular, rewrite (2.3) as

$$\left(\frac{Y - m_0(X)}{\exp(\beta_0^T X_{(1)})}\right)^2 = \exp(2g_0(X_{(2)})) + \exp(2g_0(X_{(2)}))(\varepsilon^2 - 1), \quad (2.4)$$

and estimate  $s(x_2) = \exp(2g_0(x_2))$  by kernel smoothing:

$$\hat{s}_\beta(x_2) = \frac{\sum_{i=1}^n K\left(\frac{x_2 - X_{(2)i}}{a_n}\right) \left(\frac{Y_i - \hat{m}(X_i)}{\exp(\beta^T X_{(1)i})}\right)^2}{\sum_{i=1}^n K\left(\frac{x_2 - X_{(2)i}}{a_n}\right)},$$

where  $K(\cdot)$  is a kernel function,  $a_n$  is a smoothing parameter that tends to zero as sample size gets large, and  $\hat{m}(\cdot)$  is a preliminary estimator of the regression function  $m_0(\cdot)$ .

Note that (2.4) can be rewritten as

$$(Y - m_0(X))^2 = \exp(2\beta_0^T X_{(1)})s(X_{(2)}) + \exp(2\beta_0^T X_{(1)})s(X_{(2)})(\varepsilon^2 - 1).$$

Thus we can estimate  $\beta_0$  by minimizing the following weighted least squares objective function in  $\beta$ :

$$n^{-1} \sum_{i=1}^n \frac{[(Y_i - m_0(X_i))^2 - \exp(2\beta^T X_{(1)i})s(X_{(2)i})]^2}{\exp(4\hat{\beta}^{*T} X_{(1)i})s^2(X_{(2)i})}, \quad (2.5)$$

where  $\hat{\beta}^*$  is an initial estimator of  $\beta_0$ , like e.g. the unweighted least squares estimator.

Taking the derivative of (2.5) with respect to  $\beta$ , then replacing  $s(X_{(2)i})$  and  $m_0(X_i)$  by their respective estimators, leads to the following system of equations in  $\beta$ :

$$\begin{aligned} n^{-1} \sum_{i=1}^n \left[ \frac{(Y_i - \hat{m}(X_i))^2 - \exp(2\beta^T X_{(1)i})\hat{s}_\beta(X_{(2)i})}{\exp(4\hat{\beta}^{*T} X_{(1)i})\hat{s}_{\beta^*}^2(X_{(2)i})} \right] \exp(2\beta^T X_{(1)i})\hat{s}_\beta(X_{(2)i})X_{(1)i} \\ = 0. \end{aligned} \quad (2.6)$$

Finally, the variance function can be estimated by  $\hat{\sigma}^2(x) = \exp(2\hat{\beta}^T x_1) \hat{s}_{\hat{\beta}}(x_2)$ . The above procedure can be iterated until convergence, where at each step the estimator  $\hat{\beta}^*$  is updated, and the estimated variance function is used to improve the estimator of  $m_0$ .

The estimating equation (2.6) is obtained by the backfitting method. An alternative approach is to first replace  $s(X_{(2)i})$  with the estimator  $\hat{s}_{\beta}(X_{(2)i})$  in (2.5), and then take the derivative with respect to  $\beta$ . One then also needs to take into account the dependence of  $\hat{s}_{\beta}(X_{(2)i})$  on  $\beta$ . This latter approach leads to the so-called profile estimator. We focus our attention on backfitting type estimators in this paper, see also Remark 2.1 in Section 2.3.

### 2.3. Estimation of the dispersion function with zero mean errors

We assume in this subsection that  $E(\varepsilon|X) = 0$  and  $\text{Var}(\varepsilon|X) = 1$ , in which case  $m_0(X) = E(Y|X)$  and  $\sigma^2(X, \beta_0, g_0) = \text{Var}(Y|X)$ . Similarly as in (2.3), rewrite model (2.1) as:

$$(Y - m_0(X))^2 = \sigma^2(X, \beta_0, g_0) + \sigma^2(X, \beta_0, g_0)(\varepsilon^2 - 1).$$

Then,  $\beta_0$  is the solution of the following set of equations in  $\beta$ :

$$H(\beta, g_0, m_0, w_0) = E[h(X, Y, \beta, g_0, m_0, w_0)] = 0,$$

where  $w_0(x) = \sigma^{-4}(x, \beta_0, g_0)$ ,

$$h(x, y, \beta, g, m, w) = w(x) \left[ (y - m(x))^2 - \sigma^2(x, \beta, g) \right] \frac{\partial}{\partial \beta} \sigma^2(x, \beta, g), \quad (2.7)$$

and where  $\frac{\partial}{\partial \beta} \sigma^2(x, \beta, g) = \left( \frac{\partial}{\partial \beta_j} \sigma^2(x, \beta, g(u(x, \beta))) \right)_{j=1, \dots, \ell}$ . Note that  $\frac{\partial}{\partial \beta} \sigma^2(x, \beta, g)$  is in general not only a function of  $\beta$  and  $g$ , but also of  $\frac{\partial g}{\partial u}$ , unless  $u = u(x, \beta)$  does not depend on  $\beta$ , see for example the single-index structure discussed in Section 4. The weight function  $w(x)$  belongs to some space  $\mathcal{W}$  of uniformly bounded functions and could in principle be taken as  $w(x) \equiv 1$ . However, better results are obtained in practice for the above choice of weight function, which is motivated from efficiency considerations.

Let  $\hat{m}$  be an estimator of  $m_0$ , which can be taken (in this first step) as the estimated regression function under the homoscedasticity assumption. Let  $\hat{g}(u)$  be an appropriate estimator of  $g_0(u)$  that is differentiable with respect to  $u$ . In many situations, the estimator  $\hat{g}$  depends on  $\beta$ , see for instance the motivating example in Section 2.2; we will therefore denote it by  $\hat{g}_{\beta}$  whenever the dependence on  $\beta$  is relevant. We estimate the weight  $w_0(x)$  by  $\hat{w}(x) = \sigma^{-4}(x, \hat{\beta}^*, \hat{g}_{\hat{\beta}^*})$ , where  $\hat{\beta}^*$  is (in this first step) the unweighted least squares estimator, i.e.  $\hat{\beta}^*$  satisfies  $H_n(\hat{\beta}^*, \hat{g}_{\hat{\beta}^*}, \hat{m}, 1) = 0$ , where

$$H_n(\beta, g, m, w) = n^{-1} \sum_{i=1}^n h(X_i, Y_i, \beta, g, m, w),$$

with  $h(x, y, \beta, g, m, w)$  defined in (2.7). Now, define  $\hat{\beta}$  as the solution in  $\beta$  of the equations

$$H_n(\beta, \hat{g}_\beta, \hat{m}, \hat{w}) = 0. \quad (2.8)$$

We estimate the variance function  $\sigma^2(x, \beta_0, g_0)$  by  $\hat{\sigma}^2(x) = \sigma^2(x, \hat{\beta}, \hat{g}_{\hat{\beta}})$ . This procedure can be iterated until convergence, where at each step we update the estimator  $\hat{\beta}^*$  and we re-estimate  $m_0$  by using a weighted estimation procedure that takes the heteroscedasticity into account via the estimated variance function.

**Remark 2.1.** Note that in the formula of  $H_n(\beta, \hat{g}_\beta, \hat{m}, \hat{w})$  the derivative  $\frac{\partial}{\partial \beta} \sigma^2(x, \beta, \hat{g}_\beta)$  is obtained without taking into account that  $\hat{g}_\beta$  depends on  $\beta$  (i.e. we first calculate the derivative  $\frac{\partial}{\partial \beta} \sigma^2(x, \beta, g)$  and then plug-in  $g = \hat{g}_\beta$ , thus  $\frac{\partial}{\partial \beta} \sigma^2(x, \beta, \hat{g}_\beta) = \frac{\partial}{\partial \beta} \sigma^2(x, \beta, g)|_{g=\hat{g}_\beta}$ ). As a consequence, our general estimation procedure does not cover profile estimation methods (where the derivative of  $\sigma^2(x, \beta, \hat{g}_\beta)$  takes the dependence of  $\hat{g}_\beta$  on  $\beta$  into account). However, it is easy to extend our method to profile estimators. See Section 7 for more details. For a comparison of the backfitting estimator and the profile estimator, we refer to the recent paper of Van Keilegom and Carroll (2007) and the references therein.

#### 2.4. Estimation of the dispersion function with zero median errors

Now we consider the estimation of the dispersion function when it is assumed that  $\text{med}(\varepsilon|X) = 0$  in model (2.1), which implies that  $m_0(X) = \text{med}(Y|X)$ . This can be straightforwardly extended to general quantile regression.

For identifiability of  $\sigma(x)$ , we need some additional assumption on the distribution of the random error. The assumption  $\text{med}(|\varepsilon||X) = 1$  leads to  $\sigma(X, \beta_0, g_0) = \text{med}(|Y - m_0(X)| | X)$  (median absolute deviation). An alternative common assumption is  $E(|\varepsilon| | X) = 1$ , which leads to  $\sigma(X, \beta_0, g_0) = E(|Y - m_0(X)| | X)$  (least absolute deviation). The second case is technically easier to deal with than the first. We therefore concentrate on the first case, see also Remark 3.6 in Section 3.3.

Keeping the same notations as in Section 2.3, and writing model (2.1) as

$$|Y - m_0(X)| = \sigma(X, \beta_0, g_0) + \sigma(X, \beta_0, g_0)(|\varepsilon| - 1),$$

we see that  $\beta_0$  is the solution in  $\beta$  of  $H(\beta, g_0, m_0, w_0) = E[h(X, Y, \beta, g_0, m_0, w_0)] = 0$ , where now  $w_0(x) = \sigma^{-1}(x, \beta_0, g_0)$  and

$$h(x, y, \beta, g, m, w) = w(x) \left[ 2I\{|y - m(x)| - \sigma(x, \beta, g) \geq 0\} - 1 \right] \frac{\partial}{\partial \beta} \sigma(x, \beta, g).$$

Note that the choice of the weight function is motivated from efficiency considerations. In fact, in the weighted least squares procedure that we used for the setting of zero mean errors, the weights were equal to the variance of the ‘errors’ in the ‘model’  $(Y - m_0(X))^2 = \sigma^2(X, \beta_0, g_0) + \sigma^2(X, \beta_0, g_0)(\varepsilon^2 - 1)$ . In

the current setting of zero median errors, a similar argument is used, but which is now based on the median absolute deviation of the errors instead of the mean squared deviation.

Let  $\hat{m}$  and  $\hat{g}$  be appropriate estimators of  $m_0$  and  $g_0$ , depending on the imposed model on the regression and dispersion function. Suppose that  $\hat{g}(u)$  is differentiable with respect to  $u$ . We estimate the weight function  $w_0(x)$  by  $\hat{w}(x) = \sigma^{-1}(x, \hat{\beta}^*, \hat{g}_{\hat{\beta}^*})$ , where we define the preliminary estimator  $\hat{\beta}^*$  as the solution of the non-weighted minimization problem:  $\hat{\beta}^* = \operatorname{argmin}_{\beta} \|H_n(\beta, \hat{g}_{\beta}, \hat{m}, 1)\|$ , where  $H_n(\beta, g, m, w) = n^{-1} \sum_{i=1}^n h(X_i, Y_i, \beta, g, m, w)$ , and where  $\|\cdot\|$  denotes the Euclidean norm. Finally, let

$$\hat{\beta} = \operatorname{argmin}_{\beta} \|H_n(\beta, \hat{g}_{\beta}, \hat{m}, \hat{w})\|.$$

As before, this procedure can be iterated to improve the estimation of  $\beta_0$ . Note that the function  $h$  is not smooth in  $\beta$  and hence  $\hat{\beta}$  does not necessarily satisfy  $H_n(\hat{\beta}, \hat{g}_{\hat{\beta}}, \hat{m}, \hat{w}) = 0$ .

### 3. Asymptotic results

#### 3.1. Notations and assumptions

The following notations are needed. Let  $f(y|x) = F'(y|x)$  be the density of  $Y$  given  $X = x$ , and let  $g'(u) = \frac{\partial g(u)}{\partial u}$  for any  $g \in \mathcal{G}$ . For any function  $g \in \mathcal{G}$ ,  $k \in \mathcal{K}$  and  $m \in \mathcal{M}$  (where  $\mathcal{K}$  and  $\mathcal{M}$  are the spaces to which the true functions  $g'_0$  and  $m_0$  belong respectively), we denote  $\|g\|_{\infty} = \sup_{\beta \in B} \sup_{x \in R_X} |g_{\beta}(u(x, \beta))|$ ,  $\|k\|_{\infty} = \sup_{\beta \in B} \sup_{x \in R_X} |k_{\beta}(u(x, \beta))|$  and  $\|m\|_{\infty} = \sup_{x \in R_X} |m(x)|$ . Also,  $N(\lambda, \mathcal{G}, \|\cdot\|_{\infty})$  is the covering number with respect to the norm  $\|\cdot\|_{\infty}$  of the class  $\mathcal{G}$ , i.e. the minimal number of balls of  $\|\cdot\|_{\infty}$ -radius  $\lambda$  needed to cover  $\mathcal{G}$  (see e.g. [Van der Vaart and Wellner \(1996\)](#)). Finally,  $R_U = \{u(x, \beta) : x \in R_X, \beta \in B\}$ ,  $U_0 = U(X, \beta_0)$ , and  $U_{0i} = U(X_i, \beta_0)$  ( $i = 1, \dots, n$ ).

Below we list the assumptions that are needed for the asymptotic results in Subsections 3.2 and 3.3. The purpose is to provide easy-to-check sufficient conditions such that the asymptotic results are valid for general semiparametric structures, and for both mean and median semiparametric regression models. The  $A$  and  $B$ -conditions are on the estimators  $\hat{g}$  and  $\hat{m}$  respectively, whereas all other conditions are collected under the  $C$ -list. In Section 4 we check these generic conditions for particular models and estimators of  $m_0(X)$  and  $\sigma(X, \beta_0, g_0)$ .

#### Assumptions on the estimator $\hat{g}$

- (A1)  $\|\hat{g} - g_0\|_{\infty} = o_P(1)$ ,  $\sup_{\|\beta - \beta_0\| \leq \delta_n} \sup_{x \in R_X} |(\hat{g}_{\beta} - g_{0\beta})(u(x, \beta))| = o_P(n^{-1/4})$ , and the same holds for  $\hat{g}' - g'_0$ , where  $\delta_n \rightarrow 0$ , and where  $g_{0\beta}(u)$  is such that  $g_{0\beta_0}(u) = g_0(u)$ .
- (A2)  $\int_0^{\infty} \sqrt{\log N(\lambda^s, \mathcal{H}, \|\cdot\|_{\infty})} d\lambda < \infty$ , where  $\mathcal{H} = \mathcal{G}$  or  $\mathcal{K}$ , and where  $s = 1$  for mean regression and  $s = 2$  for median regression. Moreover,  $P(\hat{g}_{\beta} \in \mathcal{G}) \rightarrow 1$  and  $P(\hat{g}'_{\beta} \in \mathcal{K}) \rightarrow 1$  as  $n$  tends to infinity, uniformly over all  $\beta$  with  $\|\beta - \beta_0\| = o(1)$ .



(A3) The estimator  $\hat{g}_0 = \hat{g}_{\beta_0}$  satisfies

$$\hat{g}_0(u) - g_0(u) = (na_n)^{-1} \sum_{i=1}^n K_1\left(\frac{u - U_{0i}}{a_n}\right) \eta(X_i, Y_i) + o_P(n^{-1/2})$$

uniformly on  $\{u(x, \beta_0) : x \in R_X\}$ , where  $E[\eta(X, Y)|X] = 0$ ,  $a_n \rightarrow 0$ ,  $na_n^{2q} \rightarrow 0$ , and  $K_1$  is a symmetric and continuous density of order  $q \geq 2$  with compact support.

(A4)  $\sup_{x \in R_X} |\{(\hat{g}_\beta - g_{0\beta}) - (\hat{g}_0 - g_0)\}(u(x, \beta))| = o_P(1)\|\beta - \beta_0\| + O_P(n^{-1/2})$ , for all  $\beta$  with  $\|\beta - \beta_0\| = o(1)$ .

**Assumptions on the estimator  $\hat{m}$**

(B1)  $\|\hat{m} - m_0\|_\infty = o_P(n^{-1/4})$ .

(B2)  $\int_0^\infty \sqrt{\log N(\lambda^s, \mathcal{M}, \|\cdot\|_\infty)} d\lambda < \infty$ , where  $s = 1$  for mean regression and  $s = 2$  for median regression. Moreover,  $P(\hat{m} \in \mathcal{M}) \rightarrow 1$  as  $n$  tends to infinity.

(B3) Uniformly in  $x \in R_X$ ,

$$\begin{aligned} \hat{m}(x) - m_0(x) &= (nb_n)^{-1} \sum_{i=1}^n \sum_{j=1}^d K_2\left(\frac{x_j - X_{ji}}{b_n}\right) \zeta_{1j}(X_{ji}, Y_i) \\ &\quad + n^{-1} \sum_{i=1}^n \zeta_2(X_i, Y_i) + o_P(n^{-1/2}), \end{aligned}$$

where  $E[\zeta_{1j}(X_j, Y)|X_j] = 0$  ( $j = 1, \dots, d$ ),  $E[\zeta_2(X, Y)] = 0$ ,  $b_n \rightarrow 0$ ,  $nb_n^4 \rightarrow 0$ , and  $K_2$  is a symmetric and continuous density with compact support.

**Other assumptions**

(C1) For all  $\delta > 0$ , there exists  $\epsilon > 0$  such that  $\inf_{\|\beta - \beta_0\| > \delta} \|H(\beta, g_0, m_0, w_0)\| \geq \epsilon > 0$ .

(C2) Uniformly for all  $\beta \in B$ ,  $H(\beta, g, m, w)$  is continuous with respect to the norm  $\|\cdot\|_\infty$  in  $(g, m, w)$  at  $(g, m, w) = (g_0, m_0, w_0)$ , and the matrix  $\Lambda$  defined in Theorem 3.1 and 3.3 is of full rank.

(C3) The function  $(x, \beta, z) \rightarrow \sigma(x, \beta, z)$  is three times continuously differentiable with respect to  $z$  and the components of  $x$  and  $\beta$ , and all derivatives are uniformly bounded on  $R_X \times B \times \{g(u) : g \in \mathcal{G}, u \in R_U\}$ . Moreover,  $\inf_{x \in R_X, \beta \in B} \sigma(x, \beta, g_0) > 0$  and  $\sup_{x, y} f(y|x) < \infty$ .

(C4) The function  $(x, \beta) \rightarrow u(x, \beta)$  is continuously differentiable with respect to the components of  $x$  and  $\beta$ , and all derivatives are uniformly bounded on  $R_X \times B$ . Moreover, the function  $(u, \beta) \rightarrow g_{0\beta}(u)$  is continuously differentiable with respect to  $u$  and the components of  $\beta$  and the derivatives are uniformly bounded on  $R_U \times B$ .

**Remark 3.1.** Let  $C_M^\alpha(R_U)$  be the set of all continuous functions  $g : R_U \rightarrow \mathbb{R}$  with

$$\|g\|_\alpha = \max_{k \leq \alpha} \sup_u |g^{(k)}(u)| + \sup_{u_1, u_2} \frac{|g^{(\alpha)}(u_1) - g^{(\alpha)}(u_2)|}{|u_1 - u_2|^{\alpha - \alpha}} \leq M < \infty,$$

where  $\underline{\alpha}$  is the largest integer strictly smaller than  $\alpha$ . Then, by Theorem 2.7.1 in Van der Vaart and Wellner (1996), the condition on the covering number in (A2) is satisfied if  $\mathcal{G}$  belongs to  $C_M^\alpha(R_U)$  with  $\alpha > 1/2$  for  $s = 1$  and  $\alpha > 1$  for  $s = 2$ .

**Remark 3.2.** In condition (A1), the function  $g_{0\beta}(u)$  satisfies  $g_{0\beta_0}(u) = g_0(u)$ . For specific examples of  $g_{0\beta}(u)$ , we refer to Section 4, particularly (4.2) and (4.4). Note that if  $u(x, \beta)$  does not depend on  $\beta$  (like for the partial linear model), then the conditions related to the derivative  $\hat{g}'$  and the space  $\mathcal{K}$  (see (A1) and (A2)) can be omitted. On the other hand, if  $u(x, \beta)$  does depend on  $\beta$ , but  $\frac{\partial}{\partial \beta} \sigma(x, \beta, g)$  is linear in  $g'(u(x, \beta))$ , then it can be easily seen that the condition  $\sup_{\|\beta - \beta_0\| \leq \delta_n} \sup_{x \in R_X} |(\hat{g}'_\beta - g'_{0\beta})(u(x, \beta))| = o_P(n^{-1/4})$  is not necessary.

**Remark 3.3.** Note that assumption (B3) requires that the regression function estimator involves at most univariate smoothing, which is the case for e.g. the partial linear, single index or additive model for the regression function, but not for the completely nonparametric model. It is possible to adapt this condition to allow for the completely nonparametric case as well, but we believe that whenever a semiparametric model is assumed for the variance function, it makes more sense to consider a semiparametric model for the regression function as well.

**Remark 3.4.** When the data are not i.i.d., the assumptions under which the main results are valid, change. In fact, these assumptions are obtained by applying the results in Chen, Linton and Van Keilegom (2003). Part of the latter paper is valid for general not necessarily i.i.d. data (namely their Theorems 1 and 2), whereas Theorem 3 is restricted to i.i.d. data. For that reason, the extension of the present paper to clustered or longitudinal data consists in replacing the assumptions that rely on Theorem 3 by corresponding assumptions valid for dependent data. The assumptions that are affected are the assumption on the bracketing number in (A2) and (B2), and assumptions (A3) and (B3), to which it should be added that the sum in the representation of  $\hat{g}_0(u)$  and  $\hat{m}(x)$  converges to a normal limit.

### 3.2. Asymptotic results with zero mean errors

In the following theorem, we give the Bahadur representation and the asymptotic normality of the estimator  $\hat{\beta}$  under the general generic conditions given in Section 3.1, and under the assumption that  $E(\varepsilon|X) = 0$  and  $\text{Var}(\varepsilon|X) = 1$ . Since  $\beta_0$  is often associated with important factors such as treatment effects, the estimation of  $\beta_0$  is sometimes of independent interest, as it tells us how the treatment affects the dispersion of the response variable in addition to its effect on the location.

The proof is given in the Appendix. We use the notation  $\frac{d}{d\beta} \sigma^2(x, \beta, g_\beta)$  to denote the complete derivative of  $\sigma^2(x, \beta, g_\beta)$  with respect to  $\beta$ , i.e.,

$$\frac{d}{d\beta}\sigma^2(x, \beta, g_\beta)_j = \lim_{\tau \rightarrow 0} \left[ \sigma^2(x, \beta + \tau e_j, g_{\beta + \tau e_j}(u(x, \beta + \tau e_j))) - \sigma^2(x, \beta, g_\beta(u(x, \beta))) \right] / \tau,$$

where  $e_j$  has the  $j$ th entry equal to one and all the other entries equal to zero,  $j = 1, \dots, \ell$ .

**Theorem 3.1.** Assume that conditions (A1)-(A4), (B1)-(B2) and (C1)-(C4) are satisfied. Then,

$$\hat{\beta} - \beta_0 = n^{-1} \sum_{i=1}^n \Lambda^{-1} \left\{ h(X_i, Y_i, \beta_0, g_0, m_0, w_0) + \xi(X_i, Y_i) \right\} + o_P(n^{-1/2}),$$

and

$$n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, \Omega),$$

where  $\Omega = \Lambda^{-1} V (\Lambda^{-1})^T$ ,

$$\begin{aligned} \Lambda &= -E \left[ \frac{1}{\sigma^4(X, \beta_0, g_0)} \frac{\partial}{\partial \beta} \sigma^2(X, \beta_0, g_0) \frac{d}{d\beta^T} \sigma^2(X, \beta_0, g_0) \right], \\ \xi(X_i, Y_i) &= -E \left[ \frac{1}{\sigma^4(X, \beta_0, g_0)} \frac{\partial}{\partial z} \sigma^2(X, \beta_0, z) \Big|_{z=g_0(U_{0i})} \right. \\ &\quad \left. \times \frac{\partial}{\partial \beta} \sigma^2(X, \beta_0, g_0) \Big| U_0 = U_{0i} \right] \eta(X_i, Y_i) f_{U_0}(U_{0i}), \\ V &= \text{Var} \left\{ h(X, Y, \beta_0, g_0, m_0, w_0) + \xi(X, Y) \right\}, \end{aligned}$$

with  $\frac{\partial}{\partial \beta} \sigma^2(x, \beta_0, g_0) = \frac{\partial}{\partial \beta} \sigma^2(x, \beta, g_0)|_{\beta=\beta_0}$ , and  $f_{U_0}(\cdot)$  the density of  $U_0$ .

Note that the above theorem does not require condition (B3). This is because the difference  $\hat{m}(X) - m_0(X)$  cancels out in the expansion of  $\hat{\beta} - \beta_0$ . As a consequence, the asymptotic variance of  $\hat{\beta} - \beta_0$  does not depend on the way we estimate the regression function  $m_0$ , since usually also the function  $\eta$  (showing up in the representation for  $\hat{g} - g_0$ ) does not depend on the way we estimate  $m_0$ . This agrees with the completely nonparametric case.

Based on the asymptotic results for  $\beta_0$ , we can establish the asymptotic normality of  $\hat{\sigma}^2(x) = \sigma^2(x, \hat{\beta}, \hat{g}_{\hat{\beta}})$ . The theorem is given below and its proof can be found in the Appendix.

**Theorem 3.2.** Assume that the conditions of Theorem 3.1 hold true. Then, for any fixed  $x \in R_X$ ,

$$(na_n)^{1/2} \left\{ \hat{\sigma}^2(x) - \sigma^2(x, \beta_0, g_0) \right\} \xrightarrow{d} N(0, v^2(x)),$$

where

$$v^2(x) = \left[ \frac{\partial}{\partial z} \sigma^2(x, \beta_0, z) \Big|_{z=g_0(u(x, \beta_0))} \right]^2 \|K_1\|_2^2 \text{Var}(\eta(X, Y) | U_0 = u(x, \beta_0)),$$

and  $\|K_1\|_2^2 = \int K_1^2(v) dv$ .

Note that the estimator  $\hat{\beta}$  does not contribute to the asymptotic variance of  $\hat{\sigma}^2(x)$ , since its rate of convergence is faster than the nonparametric rate  $(na_n)^{1/2}$ .

**Remark 3.5.** Note that the estimation of the regression function  $m_0$  can now be updated, by using a weighted least squares procedure, where the weights are given by the inverse of the estimated variance function  $\hat{\sigma}^2(x) = \sigma^2(x, \hat{\beta}, \hat{g}_{\hat{\beta}})$ . This leads to more efficient estimation of the regression function. As a special case, consider the partial linear mean regression model. Then, [Härdle, Liang and Gao \(2000\)](#) (Theorem 2.1.2, page 22) showed that whenever the estimated weights are uniformly at most  $o_P(n^{-1/4})$  away from the true (unknown) weights, then the variance of the estimators of the regression coefficients is asymptotically equal to the variance of the estimator obtained by using the true weights. In our case the weights are at a distance  $O_P((na_n)^{-1/2}) = o_P(n^{-1/4})$  away from the true weights, and so their result applies, provided we can show that this rate holds uniformly in  $x \in R_X$ . We claim that this can be shown, but the proof is long and technical and beyond the scope of this paper. Their result could be generalized to other semiparametric regression models, but we do not go deeper into this issue here (see also [Zhao \(2001\)](#) for a similar result in the context of linear median regression). It would also be of interest to consider the efficiency of the weighted least squares estimator relative to the unweighted one. We illustrate this issue in the simulation section, where we will calculate the variance of the unweighted and the weighted estimator for some specific models.

### 3.3. Asymptotic results with zero median errors

In Theorem 3.3 below, we give the Bahadur representation and the asymptotic normality of the estimator for  $\beta_0$  under the assumption that  $\text{med}(\varepsilon|X) = 0$  and  $\text{med}(|\varepsilon| |X) = 1$ . The conditional density of  $\varepsilon$  given  $X$  is denoted by  $f_\varepsilon(\cdot|X)$ .

**Theorem 3.3.** Assume that conditions (A1)-(A4), (B1)-(B3) and (C1)-(C4) are satisfied. Then,

$$\hat{\beta} - \beta_0 = n^{-1} \sum_{i=1}^n \Lambda^{-1} \left\{ h(X_i, Y_i, \beta_0, g_0, m_0, w_0) + \xi(X_i, Y_i) \right\} + o_P(n^{-1/2}),$$

and

$$n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, \Omega),$$

where  $\Omega = \Lambda^{-1} V (\Lambda^{-1})^T$ ,

$$\begin{aligned} \Lambda &= -E \left[ \frac{1}{\sigma^2(X, \beta_0, g_0)} \{ 2f_\varepsilon(1|X) + 2f_\varepsilon(-1|X) \} \frac{\partial}{\partial \beta} \sigma(X, \beta_0, g_0) \frac{d}{d\beta^T} \sigma(X, \beta_0, g_0) \right], \\ &\quad \xi(X_i, Y_i) \\ &= \sum_{j=1}^d E \left[ \frac{1}{\sigma^2(X, \beta_0, g_0)} \{ -2f_\varepsilon(1|X) + 2f_\varepsilon(-1|X) \} \right] \end{aligned}$$

$$\begin{aligned}
 & \times \frac{\partial}{\partial \beta} \sigma(X, \beta_0, g_0) \Big| X_j = X_{ji} \Big] \zeta_{1j}(X_{ji}, Y_i) f_{X_j}(X_{ji}) \\
 & + E \left[ \frac{1}{\sigma^2(X, \beta_0, g_0)} \{ -2f_\varepsilon(1|X) + 2f_\varepsilon(-1|X) \} \frac{\partial}{\partial \beta} \sigma(X, \beta_0, g_0) \right] \zeta_2(X_i, Y_i) \\
 & - E \left[ \frac{1}{\sigma^2(X, \beta_0, g_0)} \{ 2f_\varepsilon(1|X) + 2f_\varepsilon(-1|X) \} \frac{\partial}{\partial z} \sigma(X, \beta_0, z) \Big|_{z=g_0(U_{0i})} \right. \\
 & \quad \left. \times \frac{\partial}{\partial \beta} \sigma(X, \beta_0, g_0) \Big| U_0 = U_{0i} \right] \eta(X_i, Y_i) f_{U_0}(U_{0i}), \\
 V = & \text{Var} \left\{ h(X, Y, \beta_0, g_0, m_0, w_0) + \xi(X, Y) \right\}.
 \end{aligned}$$

**Theorem 3.4.** Assume that the conditions of Theorem 3.3 hold true. Then, for any fixed  $x \in R_X$ ,

$$(na_n)^{1/2} \left\{ \hat{\sigma}(x) - \sigma(x, \beta_0, g_0) \right\} \xrightarrow{d} N(0, v^2(x)),$$

where

$$v^2(x) = \left[ \frac{\partial}{\partial z} \sigma(x, \beta_0, z) \Big|_{z=g_0(u(x, \beta_0))} \right]^2 \|K_1\|_2^2 \text{Var}(\eta(X, Y) | U_0 = u(x, \beta_0)).$$

**Remark 3.6.** The above two theorems can be easily adapted to the case where the dispersion function is defined by  $\sigma(x, \beta, g) = E(|Y - m(X)| | X = x)$  (i.e.  $E(|\varepsilon| | X) = 1$ ). In fact, the formulas of the matrix  $\Lambda$  and of the function  $\xi$  can be similarly obtained by combining the calculations done in the proofs of Theorems 3.1 and 3.3. These calculations show that the parameter  $s$  in condition (B2) equals 2, whereas for condition (A2)  $s$  equals 1. We omit the details.

## 4. Examples

In this section we consider two particular semiparametric regression models, we propose estimators under these models and verify the conditions that are required for the asymptotic results of Section 3. The first example is a representative example for mean regression, the second one for median regression.

### 4.1. Single index mean regression model

In this first example we consider a mean regression model with a single index regression and variance function:

$$Y = r_0(\alpha_0^T X) + g_0^{1/2}(\beta_0^T X) \varepsilon, \quad (4.1)$$

where  $E(\varepsilon|X) = 0$ ,  $E(\varepsilon^2|X) = 1$  and where  $g_0$  is a positive function. In order to correctly identify the model, we assume that  $\alpha_{01} = \beta_{01} = 1$ . This model has also been studied by Xia, Tong and Li (2002), using a different estimation method.

Let  $\hat{m}(x)$  be an estimator of the unknown regression function  $m_0(x) = r_0(\alpha_0^T x)$ , like e.g. the estimator proposed in [Härdle, Hall and Ichimura \(1993\)](#). See also [Delecroix, Hristache and Patilea \(2006\)](#) for a more general class of semiparametric  $M$ -estimators of  $m_0(x)$ . Since the verification of conditions (B1) and (B2) is easier than of conditions (A1) and (A2), we concentrate in what follows on the verification of the A-conditions. First, define for any  $\beta \in \mathbb{R}^d$ ,

$$g_{0\beta}(u) = E\left((Y - m_0(X))^2 | \beta^T X = u\right), \quad (4.2)$$

and let

$$\hat{g}_\beta(u) = \sum_{i=1}^n \frac{K_{1a}(u - \beta^T X_i)}{\sum_{j=1}^n K_{1a}(u - \beta^T X_j)} (Y_i - \hat{m}(X_i))^2,$$

where  $K_{1a}(v) = K_1(v/a_n)/a_n$ ,  $K_1$  is a kernel function and  $a_n$  a bandwidth sequence. For (A1), note that

$$\begin{aligned} \hat{g}_\beta(u) - g_{0\beta}(u) &= \sum_{i=1}^n \frac{K_{1a}(u - \beta^T X_i)}{\sum_{j=1}^n K_{1a}(u - \beta^T X_j)} (Y_i - m_0(X_i))^2 - g_{0\beta}(u) \\ &\quad + \sum_{i=1}^n \frac{K_{1a}(u - \beta^T X_i)}{\sum_{j=1}^n K_{1a}(u - \beta^T X_j)} \left\{ (\hat{m}(X_i) - m_0(X_i))^2 \right. \\ &\quad \left. - 2(Y_i - m_0(X_i))(\hat{m}(X_i) - m_0(X_i)) \right\} \\ &= O_P((na_n)^{-1/2}(\log n)^{1/2}) + o_P(n^{-1/4}) = o_P(n^{-1/4}), \end{aligned}$$

uniformly in  $u$  and  $\beta$ , provided  $na_n^2(\log n)^{-2} \rightarrow \infty$  and  $\inf_{\beta \in B} \inf_{x \in R_X} f_{\beta^T X} \times (\beta^T x) > 0$  (where  $f_{\beta^T X}$  is the density of  $\beta^T X$ ). For  $\hat{g}'_\beta$ , note that  $\frac{\partial}{\partial \beta} \sigma^2(x, \beta, g) = g'(\beta^T x)x$  is linear in  $g'(\beta^T x)$ , and hence, by [Remark 3.2](#), we only need to show that  $\|\hat{g}' - g'_0\|_\infty = o_P(1)$ . This can be shown using standard calculations. Next, let  $\mathcal{G} = \mathcal{K} = C_M^{1/2+\delta}(R_U)$  for some  $\delta > 0$ . It follows from [Remark 3.1](#) that the condition on the covering number of  $\mathcal{G}$  and  $\mathcal{K}$  in (A2) is satisfied. Moreover,  $\sup_{u,\beta} |\hat{g}_\beta(u) - g_{0\beta}(u)| = \sup_{u,\beta} |\hat{g}_\beta(u)| + o_P(1) = O_P(1)$  (and similarly for  $\hat{g}'_\beta(u)$ ), and  $\sup_{\beta, u_1, u_2} |\hat{g}'_\beta(u_1) - \hat{g}'_\beta(u_2)| / |u_1 - u_2|^{1/2+\delta} \leq M$  provided  $na_n^{4+2\delta}(\log n)^{-1} \rightarrow \infty$ . Hence,  $P(\hat{g}_\beta \in \mathcal{G}) \rightarrow 1$  and  $P(\hat{g}'_\beta \in \mathcal{K}) \rightarrow 1$ . For (A3), note that

$$\begin{aligned} \hat{g}_0(u) - g_0(u) &= \sum_{i=1}^n \frac{K_{1a}(u - \beta_0^T X_i)}{\sum_{j=1}^n K_{1a}(u - \beta_0^T X_j)} (Y_i - m_0(X_i))^2 - g_0(u) \\ &\quad - 2 \sum_{i=1}^n \frac{K_{1a}(u - \beta_0^T X_i)}{\sum_{j=1}^n K_{1a}(u - \beta_0^T X_j)} (Y_i - m_0(X_i))(\hat{m}(X_i) - m_0(X_i)) + o_P(n^{-1/2}). \end{aligned}$$

Let  $K_1$  be a kernel of order  $q \geq 3$ . Then, the first term above can be written as

$$n^{-1} \sum_{i=1}^n K_{1a}(u - \beta_0^T X_i) \left\{ (Y_i - m_0(X_i))^2 - g_0(\beta_0^T X_i) \right\} f_{\beta_0^T X}^{-1}(\beta_0^T X_i) + o(n^{-1/2}),$$

provided  $na_n^6 \rightarrow 0$ . The second term is a degenerate  $V$ -process (with kernel depending on  $n$ ), and can be written as a degenerate  $U$ -process, plus a term of order  $O_P((nb_n)^{-1}) = o_P(n^{-1/2})$  provided  $nb_n^2 \rightarrow \infty$ . The  $U$ -process can be written out using Hajek-projection techniques, similar to the ones for regular degenerate  $U$ -statistics, which shows at the end (after long but straightforward calculations) that this term is  $o_P(n^{-1/2})$  provided  $na_nb_n \rightarrow \infty$ . Hence, (A3) holds true for  $\eta(x, y) = \{(y - m_0(x))^2 - g_0(\beta_0^T x)\} f_{\beta_0^T X}^{-1}(\beta_0^T x)$ . Finally, for (A4),

$$|(\hat{g}_\beta - g_{0\beta} - \hat{g}_0 + g_0)(\beta^T x)| \leq \left| \frac{\partial}{\partial \beta} [\hat{g}_{\tilde{\beta}} - g_{0\tilde{\beta}}](\beta^T x) (\beta - \beta_0) \right| = o_P(1) \|\beta - \beta_0\|,$$

uniformly in  $\beta$  and  $x$ , where  $\tilde{\beta}$  is between  $\beta_0$  and  $\beta$ . It now follows that  $\hat{\beta} - \beta_0$  is asymptotically normal, with mean zero and variance given in Theorem 3.1.

#### 4.2. Partially linear median regression model

The second model we consider is a median regression model with a partially linear regression function and an exponentially transformed partially linear dispersion function :

$$Y = \alpha_0^T X_{(1)} + r_0(X_{(2)}) + \exp(\beta_0^T X_{(1)} + g_0(X_{(2)}))\varepsilon, \quad (4.3)$$

where  $\text{med}(\varepsilon|X) = 0$ ,  $E(|\varepsilon||X) = 1$ , and  $X = (X_{(1)}^T, X_{(2)})^T$ , with  $X_{(1)} = (X_1, \dots, X_{d-1})^T$  and  $X_{(2)} = X_d$ . For any  $\beta \in \mathbb{R}^{d-1}$  and for  $m_0(x) = \alpha_0^T x_1 + r_0(x_2)$ , let  $g_{0\beta}(x_2) = \log s_{0\beta}(x_2)$ , and  $\hat{g}_\beta(x_2) = \log \hat{s}_\beta(x_2)$ , where

$$s_{0\beta}(x_2) = E\left(\frac{|Y - m_0(X)|}{\exp(\beta^T X_{(1)})} \mid X_{(2)} = x_2\right), \quad (4.4)$$

and

$$\hat{s}_\beta(x_2) = \sum_{i=1}^n \frac{K_{1a}(x_2 - X_{(2)i})}{\sum_{j=1}^n K_{1a}(x_2 - X_{(2)j})} \frac{|Y_i - \hat{m}(X_i)|}{\exp(\beta^T X_{(1)i})},$$

where  $\hat{m}(x) = \hat{\alpha}^T x_1 + \hat{r}(x_2)$  is an estimator of the unknown regression function  $m_0(x)$ , see e.g. Härdle, Liang and Gao (2000, Chapter 2). Define

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} n^{-1} \sum_{i=1}^n \left\{ \frac{|Y_i - \hat{m}(X_i)| - \exp(\beta^T X_{(1)i}) \hat{s}_\beta(X_{(2)i})}{\exp(\hat{\beta}^{*T} X_{(1)i}) \hat{s}_{\hat{\beta}^*}(X_{(2)i})} \right\}^2,$$

where

$$\hat{\beta}^* = \underset{\beta}{\operatorname{argmin}} n^{-1} \sum_{i=1}^n \left\{ |Y_i - \hat{m}(X_i)| - \exp(\beta^T X_{(1)i}) \hat{s}_\beta(X_{(2)i}) \right\}^2.$$

As in the previous example, we restrict attention to verifying the A-conditions. Since  $u(X, \beta) = X_{(2)}$  does not depend on  $\beta$ , we do not need to check the conditions related to  $\hat{g}'$  and  $\mathcal{K}$ . Note that

$$|\hat{s}_\beta(x_2) - s_{0\beta}(x_2)| \leq \left| \sum_{i=1}^n \frac{K_{1a}(x_2 - X_{(2)i})}{\sum_{j=1}^n K_{1a}(x_2 - X_{(2)j})} \frac{|Y_i - m_0(X_i)|}{\exp(\beta^T X_{(1)i})} - s_{0\beta}(x_2) \right|$$

$$\begin{aligned}
& + \sum_{i=1}^n \frac{K_{1a}(x_2 - X_{(2)i})}{\sum_{j=1}^n K_{1a}(x_2 - X_{(2)j})} |\hat{m}(X_i) - m_0(X_i)| \\
& = O_P((na_n)^{-1/2}(\log n)^{1/2}) + o_P(n^{-1/4}) = o_P(n^{-1/4})
\end{aligned}$$

uniformly in  $x$  and  $\beta$  if  $na_n^2(\log n)^{-2} \rightarrow \infty$ . Hence, (A1) is satisfied, provided  $\inf_{x_2, \beta} s_{0\beta}(x_2) > 0$ . For (A2) similar arguments as in the first example show that  $\mathcal{G} = C_M^1(R_{X_{(2)}})$  can be used. Next, consider the verification of condition (A3). Using the property that for any  $x, y$ ,

$$|x - y| - |x| = 2(-y)\psi(x) + 2(y - x)[I(y > x > 0) - I(y < x < 0)],$$

where  $\psi(x) = 0.5 - I(x < 0)$ , we have

$$\begin{aligned}
& \hat{s}_0(x_2) - s_0(x_2) \\
& = \sum_{i=1}^n \frac{K_{1a}(x_2 - X_{(2)i})}{\sum_{j=1}^n K_{1a}(x_2 - X_{(2)j})} \frac{|Y_i - \hat{m}(X_i)|}{\exp(\beta_0^T X_{(1)i})} - s_0(x_2) \\
& = \sum_{i=1}^n \frac{K_{1a}(x_2 - X_{(2)i})}{\sum_{j=1}^n K_{1a}(x_2 - X_{(2)j})} \exp(-\beta_0^T X_{(1)i}) \left\{ |Y_i - m_0(X_i)| \right. \\
& \quad - 2(\hat{m}(X_i) - m_0(X_i))\psi(Y_i - m_0(X_i)) \\
& \quad - 2(Y_i - \hat{m}(X_i))[I(\hat{m}(X_i) - m_0(X_i) > Y_i - m_0(X_i) > 0) \\
& \quad \left. - I(\hat{m}(X_i) - m_0(X_i) < Y_i - m_0(X_i) < 0)] \right\} - s_0(x_2) \\
& = A(x_2) + B(x_2) + C(x_2) - s_0(x_2) \quad (\text{say}).
\end{aligned}$$

First consider

$$\begin{aligned}
A(x_2) - s_0(x_2) & = n^{-1} \sum_{i=1}^n K_{1a}(x_2 - X_{(2)i}) \left\{ \frac{|Y_i - m_0(X_i)|}{\exp(\beta_0^T X_{(1)i})} - s_0(X_{(2)i}) \right\} \\
& \quad \times f_{X_{(2)}}^{-1}(X_{(2)i}) + o_P(n^{-1/2}),
\end{aligned}$$

provided  $na_n^4 \rightarrow 0$  and  $K_1$  is a kernel of order 2. Next, note that the term  $B(x_2)$  is a degenerate  $V$ -process, because in the i.i.d. representation of  $\hat{m}(X_i) - m_0(X_i)$ , each term has mean zero, and because  $E(\psi(Y_i - m_0(X_i))|X_i) = 0$ . Hence, as for the first example, we have that  $B(x_2) = o_P(n^{-1/2})$ . Finally, using the notation  $\epsilon_i = Y_i - m_0(X_i)$  and  $\hat{d} = \hat{m} - m_0$ , consider

$$\begin{aligned}
& |C(x_2)| \\
& \leq 2 \sum_{i=1}^n \frac{K_{1a}(x_2 - X_{(2)i})}{\sum_{j=1}^n K_{1a}(x_2 - X_{(2)j})} \exp(-\beta_0^T X_{(1)i}) \left( |\epsilon_i| + |\hat{d}(X_i)| \right) I\left(|\epsilon_i| < |\hat{d}(X_i)|\right) \\
& \leq 4 \sum_{i=1}^n \frac{K_{1a}(x_2 - X_{(2)i})}{\sum_{j=1}^n K_{1a}(x_2 - X_{(2)j})} \exp(-\beta_0^T X_{(1)i}) |\hat{d}(X_i)| I\left(|\epsilon_i| < |\hat{d}(X_i)|\right)
\end{aligned}$$



$$\begin{aligned} &\leq 4 \sup_{x_1} \exp(-\beta_0^T x_1) \sup_x |\hat{d}(x)| \left( \inf_{x_2} f_{X_{(2)}}(x_2) \right)^{-1} n^{-1} \\ &\quad \times \sum_{i=1}^n K_{1a}(x_2 - X_{(2)i}) I(|\epsilon_i| < |\hat{d}(X_i)|) + o_P(n^{-1/2}), \end{aligned}$$

since  $\sup_x |\hat{d}(x)| = o_P(n^{-1/4})$  by (B1). Using e.g. Van der Vaart and Wellner (1996, Section 2.11), it can be shown that the process

$$n^{-1} \sum_{i=1}^n K_{1a}(x_2 - X_{(2)i}) I(|\epsilon_i| < |d(X_i)|) - P(|\epsilon| \leq |d(X)| | X_{(2)} = x_2)$$

is  $O_P((na_n)^{-1/2}(\log n)^{1/2})$  uniformly in  $x_2 \in R_{X_{(2)}}$  and in  $d = m - m_0$  with  $m \in \mathcal{M}$ . Hence,

$$\begin{aligned} &n^{-1} \sum_{i=1}^n K_{1a}(x_2 - X_{(2)i}) I(|\epsilon_i| < |\hat{d}(X_i)|) \\ &= P(|\epsilon| \leq |\hat{d}(X)| | X_{(2)} = x_2) + O_P((na_n)^{-1/2}(\log n)^{1/2}), \end{aligned}$$

since  $P(\hat{m} \in \mathcal{M}) \rightarrow 1$  by condition (B2). It now follows that

$$\begin{aligned} |C(x_2)| &= O\left(\sup_x |\hat{d}(x)|\right) \left\{ P(|\epsilon| \leq |\hat{d}(X)| | X_{(2)} = x_2) \right. \\ &\quad \left. + O_P((na_n)^{-1/2}(\log n)^{1/2}) \right\} \\ &= o_P(n^{-1/4}) \int [F_\epsilon(|\hat{d}(x)| | x) - F_\epsilon(-|\hat{d}(x)| | x)] f_{X_{(1)}}(x_1) dx_1 + o_P(n^{-1/2}) \\ &= o_P(n^{-1/4}) 2 \sup_{x,y} f_\epsilon(y|x) \sup_x |\hat{d}(x)| + o_P(n^{-1/2}) \\ &= o_P(n^{-1/2}). \end{aligned}$$

This finishes the proof for condition (A3). It remains to check (A4), which can be done in much the same way as in the first example.

## 5. A Monte-Carlo example

We generate random data from a partially linear heteroscedastic median regression model given by

$$Y_i = \alpha_0^T X_{(1)i} + r_0(X_{2i}) + \exp(\beta_0 X_{1i} + g_0(X_{2i})) \varepsilon_i, \quad i = 1, \dots, 400,$$

where  $X_i = (X_{(1)i}^T, X_{2i})^T$ ,  $X_{(1)i} = (1, X_{1i})^T$ ,  $\alpha_0 = (\alpha_{00}, \alpha_{01})^T = (1, 0.75)^T$ ,  $\beta_0 = -2$ ,  $X_{1i}$  and  $X_{2i}$  are independent with uniform distribution on  $(0,1)$ , and  $\varepsilon_i$  ( $i = 1, \dots, n$ ) are i.i.d. normal random variables that are independent of  $X_{1i}$  and  $X_{2i}$ . Furthermore, the  $\varepsilon_i$  are standardized such that  $\text{med}(\varepsilon_i) = 0$

and  $E(|\varepsilon_i|) = 1$ . We consider the following choices for  $r_0(x_2)$  and  $g_0(x_2)$ , where  $r_0(X_2)$  satisfies the zero median constraint in order for the intercept to be identifiable.

- Case 1:  $r_0(x_2) = 2(x_2 - 0.5)$  and  $g_0(x_2) = 1.4 - 2x_2$ ;  
Case 2:  $r_0(x_2) = 2(x_2 - 0.5)$  and  $g_0(x_2) = 1.1 - 2x_2^2$ ;  
Case 3:  $r_0(x_2) = \exp(-x_2) - \exp(-0.5)$  and  $g_0(x_2) = 1.4 - 2x_2$ ;  
Case 4:  $r_0(x_2) = \exp(-x_2) - \exp(-0.5)$  and  $g_0(x_2) = 1.1 - 2x_2^2$ .  
Case 5:  $r_0(x_2) = \sin(4(x_2 - 0.5))$  and  $g_0(x_2) = 1.4 - 2x_2$ .  
Case 6:  $r_0(x_2) = \sin(4(x_2 - 0.5))$  and  $g_0(x_2) = 1.1 - 2x_2^2$ .

The model is fitted using the backfitting algorithm described in Section 2.4. Implementation of the algorithm involves two smoothing parameters, one for estimating the conditional median function and the other for estimating the dispersion function. For the former, we apply the automatic smoothing parameter selection method in Yu and Jones (1998), while for the latter, we consider smoothing parameters on the grid  $[0.02, 0.20]$  with step size 0.02. We choose the latter smoothing parameter by cross-validation such that the ability to estimating the conditional median function is optimized. This algorithm works well in our simulations. We have not encountered any convergence problem. Alternatively, one may apply the cross-validation method to the two smoothing parameters jointly. This is, however, much more computationally intensive.

We report results from 500 independent simulation runs. First, we compare the unweighted method with the weighted method for estimating  $\alpha_0$ . The unweighted method assumes that the dispersion function is constant; while the weighted method updates the estimator  $\hat{\alpha}$  via the weighted  $L_1$  regression where the weights are taken to be the reciprocal of the estimated dispersion function. Table 1 displays the bias and the MSE for estimating  $\alpha_{00}$  and  $\alpha_{01}$ , respectively. It also reports the *simulated relative efficiency* (SRE) for comparing the weighted and unweighted methods. The SRE is defined as

$$\text{SRE} = \frac{\text{MSE for estimating } \alpha_0 \text{ using the unweighted method}}{\text{MSE for estimating } \alpha_0 \text{ using the weighted method}},$$

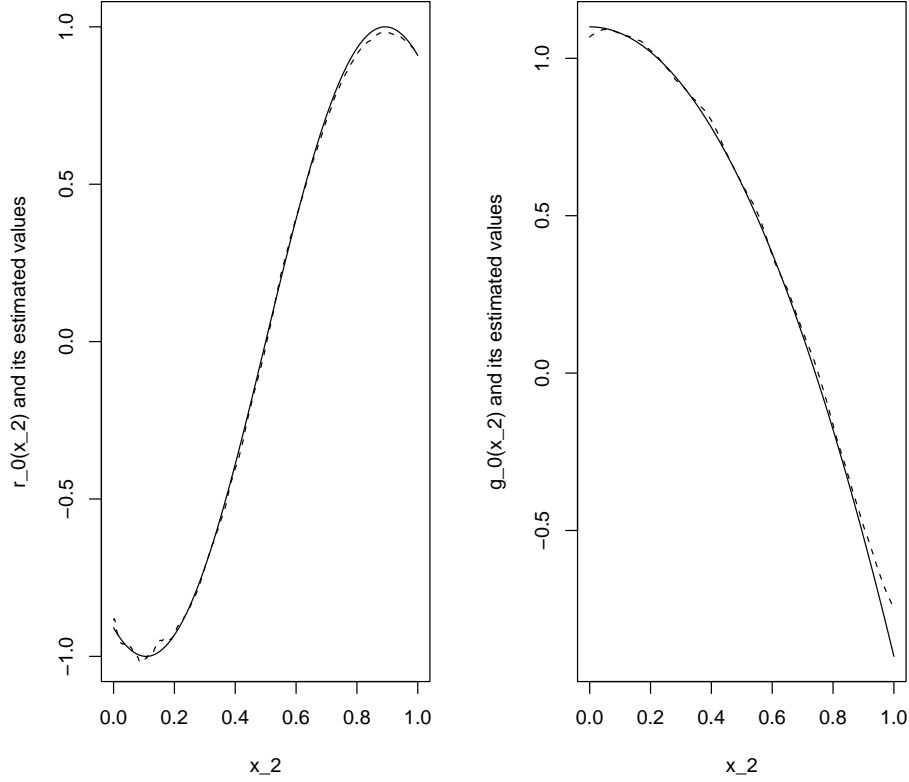
where the MSE for estimating  $\alpha_0$  is defined as the sum of the mean squared errors for estimating each coordinate of  $\alpha_0$ . The simulation results suggest that the weighted method significantly improves the efficiency of estimating  $\alpha_0$  compared with the unweighted method. In all six cases, we observe an efficiency gain around 20-30% when using the weighted method.

TABLE 1  
Estimating  $\alpha_0$ : comparing unweighted and weighted methods

	$\hat{\alpha}_{00}$ (unweighted)		$\hat{\alpha}_{01}$ (unweighted)		$\hat{\alpha}_{00}$ (weighted)		$\hat{\alpha}_{01}$ (weighted)		SRE for estimating $\alpha_0$
	bias	MSE	bias	MSE	bias	MSE	bias	MSE	
Case 1	0.051	0.072	-0.065	0.079	0.035	0.062	-0.040	0.060	1.238
Case 2	0.048	0.075	-0.059	0.076	0.031	0.064	-0.035	0.056	1.258
Case 3	0.009	0.053	-0.046	0.072	-0.010	0.042	-0.019	0.053	1.312
Case 4	0.021	0.057	-0.059	0.076	0.005	0.047	-0.035	0.056	1.291
Case 5	0.063	0.085	-0.071	0.075	0.040	0.068	-0.030	0.054	1.305
Case 6	0.031	0.072	-0.050	0.078	0.013	0.061	-0.021	0.052	1.329

TABLE 2  
Estimating  $\beta_0$ 

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
bias	-0.052	-0.049	-0.040	-0.048	-0.038	-0.059
MSE	0.024	0.025	0.022	0.025	0.023	0.028

FIG 1. Estimates of the functions  $r_0(x_2)$  and  $g_0(x_2)$  for case 6. Solid line: true curve; dashed line: estimated curve.

Next, we consider estimating the dispersion parameter  $\beta_0$  when the weighted method is used. Table 2 gives the bias and the mean squared error, which suggests that  $\beta_0$  is estimated satisfactorily in all cases.

Finally, we give some idea on how well we estimate the nonparametric parts of the semiparametric model. More specifically, we consider case 6 and compare in Figure 1 the true curves of  $r_0(x_2)$  and  $g_0(x_2)$  with their respective estimates (averaged over the 500 simulation runs). The estimated curves are very close to the true curves. Results from all other cases are similar and not reported due to space limitation.

## 6. Analysis of gasoline consumption data

We illustrate the proposed method by means of a data set on gasoline consumption. The data were collected by the National Private Vehicle Use Survey in Canada between October 1994 and September 1996 and contain household-based information (Yatchew, 2003). In this analysis, we use the subset of September data which consists of 485 observations. We are interested in estimating the median of the log of the distance traveled per month by the household (denoted by  $Y = \text{dist}$ ) based on six covariates:  $X_1 = \text{income}$  (log of the previous year's combined annual household income before taxes which is reported in 9 ranges),  $X_2 = \text{driver}$  (log of the number of the licensed drivers in the household),  $X_3 = \text{age}$  (log of the age of driver),  $X_4 = \text{retire}$  (a dummy variable for those households whose head is over the age of 65),  $X_5 = \text{urban}$  (a dummy variable for urban dwellers), and  $X_6 = \text{price}$  (log of the price of a liter of gasoline). The scatter plots of the response variable versus each covariate are given in Figure 2.

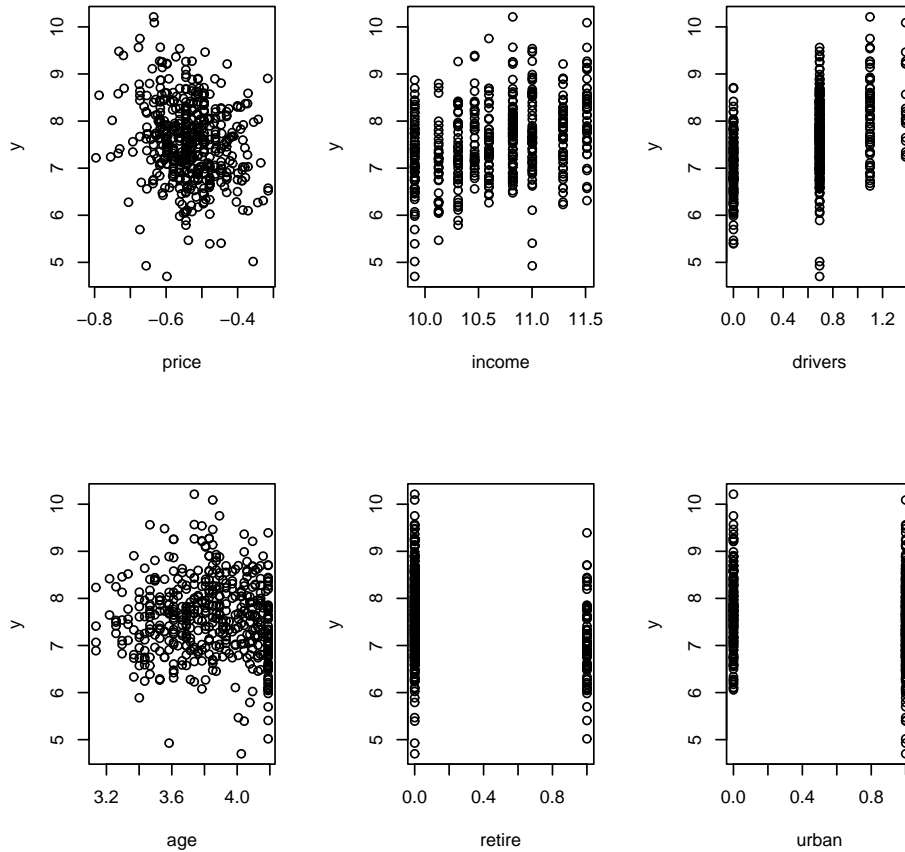


FIG 2. Plot for the gasoline consumption data.

TABLE 3  
Analysis of the gasoline consumption data (*sd* = standard error)

	$\alpha_0$		$\beta_0$	
	estimate	sd	estimate	sd
<i>income</i>	0.351	0.099	-0.009	0.046
<i>driver</i>	0.662	0.087	0.245	0.108
<i>age</i>	0.193	0.162	-0.037	0.126
<i>retire</i>	-0.283	0.130	0.169	0.129
<i>urban</i>	-0.288	0.080	0.097	0.074

We fit a heteroscedastic partially linear median regression model, which was motivated by a homoscedastic partially linear mean regression model by [Yatchew \(2003\)](#). More specifically, we assume

$$\begin{aligned} dist = & \alpha_{01}income + \alpha_{02}driver + \alpha_{03}age + \alpha_{04}retire + \alpha_{05}urban + r_0(price) \\ & + \exp[\beta_{01}income + \beta_{02}driver + \beta_{03}age + \beta_{04}retire \\ & + \beta_{05}urban + g_0(price)]\varepsilon, \end{aligned}$$

i.e.  $Y = \alpha_0^T X_{(1)} + r_0(X_{(2)}) + \exp(\beta_0^T X_{(1)} + g_0(X_{(2)}))\varepsilon$ , where  $X = (X_{(1)}^T, X_{(2)}^T)^T$  with  $X_{(1)} = (X_1, \dots, X_5)^T$  and  $X_{(2)} = X_6 = price$ , and where  $r_0(\cdot)$  and  $g_0(\cdot)$  are two unknown smooth functions. For identifying the model we assume that  $\text{med}(\varepsilon|X) = 0$  and  $E(|\varepsilon||X) = 1$ . The smoothing parameters are selected using the approach described in Section 5. The smoothing parameter for estimating the conditional median function is 0.02 and that for estimating the dispersion function is 0.05.

Table 3 summarizes the estimated coefficients in the parametric parts of the conditional median function and the dispersion function. It is not surprising that households with larger income and more drivers tend to have higher median value of *dist*, and that retired people and urban dwellers tend to drive less. Table 3 also contains the standard errors of the  $\hat{\alpha}_j$ 's and  $\hat{\beta}_j$ 's. These are obtained using a model-based resampling procedure. More specifically, we estimate the parametric and nonparametric components in the above model and obtain  $\hat{\alpha}$ ,  $\hat{\beta}$ ,  $\hat{r}$  and  $\hat{g}$ . We then generate a bootstrap sample ( $i = 1, \dots, n$ ):  $Y_i^* = \sum_{j=1}^5 \hat{\alpha}_j X_{ji} + \hat{r}(X_{6i}) + \exp(\sum_{j=1}^5 \hat{\beta}_j X_{ji} + \hat{g}(X_{6i}))\varepsilon_i^*$ , where the  $\varepsilon_i^*$  satisfy the constraints  $\text{med}(\varepsilon_i^*|X_i) = 0$  and  $E(|\varepsilon_i^*||X_i) = 1$  (we use a normal distribution in the simulations). For each bootstrap sample, we re-estimate the  $\alpha_j$ 's and the  $\beta_j$ 's. The standard errors are then calculated from these estimators based on 200 bootstrap samples. The results in Table 3 suggest that *income*, *driver*, *retire* and *urban* have significant effects on the conditional median function. Moreover, *driver* exhibits a significant effect on the dispersion function.

Figure 3 displays the estimated nonparametric components. The plots indicate that for the majority of values of *price*, increased *price* is associated with reduced conditional median of *dist*. However, for the lowest and highest values of *price*, the effect of *price* on *dist* seems to be reversed. A similar pattern is observed for the effect of *price* on the dispersion function. Note however that for small and large values of *price*, the data are rather sparse, as can be seen from Figure 2.

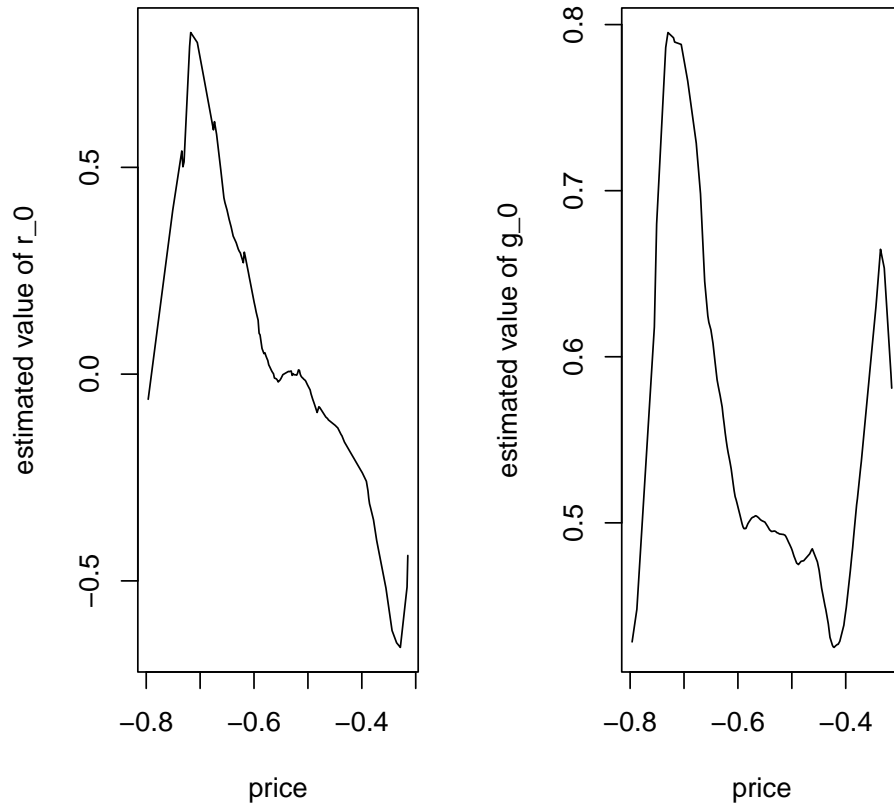


FIG 3. Estimates of the functions  $r_0(\text{price})$  and  $g_0(\text{price})$  for the gasoline consumption data.

## 7. Discussion

This paper considers a general class of semiparametric location-dispersion models. The theory we have developed focuses on how to estimate the dispersion function and the theoretical properties of the proposed estimators.

The estimators we use are of the back-fitting type. Alternatively, one may consider profile estimators, which are obtained by replacing in the definition of  $h(x, y, \beta, g_\beta, m, w)$  the partial derivative  $\frac{\partial}{\partial \beta} \sigma^2(x, \beta, g_\beta)$  by the complete derivative  $\frac{d}{d\beta} \sigma^2(x, \beta, g_\beta)$ , i.e. profile estimators take into account that  $g_\beta$  also depends on  $\beta$ . See [Van Keilegom and Carroll \(2007\)](#) for a detailed analysis of the pros and cons of profiling versus backfitting.

In the future, we would like to study in more detail the estimation of the mean or median using weighted least squares with weights equal to the inverse of the estimated dispersion function. The simulations in [Section 5](#) suggest that the efficiency gain is quite substantial. A theoretical analysis of the relative efficiency will be very interesting. Certainly, we would also like to extend this class of models to time-series setting.

## Appendix: Proofs

*Proof of Theorem 3.1.* We will make use of Theorem 2 in Chen, Linton and Van Keilegom (2003) (CLV hereafter), which gives generic conditions under which  $\hat{\beta}$  is asymptotically normal. First of all, we need to show that  $\hat{\beta} - \beta_0 = o_P(1)$ . For this, we verify the conditions of Theorem 1 in CLV. Condition (1.1) holds by definition of  $\hat{\beta}$ , while the second and third condition are guaranteed by assumptions (C1) and (C2). For condition (1.4) we use assumptions (A1) and (B1) for the estimators  $\hat{g}$  and  $\hat{m}$ , whereas for  $\hat{w}$  more work is needed. In fact, one should first consider the current proof in the case where  $w \equiv 1$ . In that case the function  $w$  is not a nuisance function and redoing the whole proof with  $w \equiv 1$ , we get at the end that  $\hat{\beta}^* - \beta_0 = O_P(n^{-1/2})$  and  $\|\hat{w} - w_0\|_\infty = o_P(n^{-1/4})$ . Finally, condition (1.5) is very similar to condition (2.5) of Theorem 2 of CLV, and we will verify both conditions below. So, the conditions of Theorem 1 are verified, up to condition (1.5) which we postpone to later. Next, we verify conditions (2.1)–(2.6) of Theorem 2 in CLV. Condition (2.1) is, as for condition (1.1), valid by construction of the estimator  $\hat{\beta}$ . For condition (2.2), first note that since  $U = U(X, \beta)$  depends in general on  $\beta$ , the criterion function  $h$  does not only depend on the nuisance functions  $g_\beta, m$  and  $w$ , but also on  $g'_\beta(u) = \frac{\partial g_\beta(u)}{\partial u}$ . Therefore, from now on we will denote  $H(\beta, g_\beta, m, w, g'_\beta)$  to stress the dependence on  $g'_\beta$ . Similar, whenever it is necessary to stress the dependence of  $\frac{\partial}{\partial \beta} \sigma^2(x, \beta, g_\beta)$  on  $g'_\beta$ , we will write  $\sigma_\beta^2(x, \beta, g_\beta, g'_\beta)$ . We need to calculate the matrix

$$\begin{aligned} & \frac{d}{d\beta^T} H(\beta, g_{0\beta}, m_0, w_0, g'_{0\beta}) \\ &= E \left[ -w_0(X) \frac{\partial}{\partial \beta} \sigma^2(X, \beta, g_{0\beta}) \frac{d}{d\beta^T} \sigma^2(X, \beta, g_{0\beta}) \right. \\ & \quad \left. + w_0(X) \left\{ \sigma^2(X, \beta_0, g_0) - \sigma^2(X, \beta, g_{0\beta}) \right\} \frac{d}{d\beta^T} \frac{\partial}{\partial \beta} \sigma^2(X, \beta, g_{0\beta}) \right]. \end{aligned}$$

Note that when  $\beta = \beta_0$ , the second term above equals zero and we find the matrix  $\Lambda$  in that case. Hence, (2.2) follows from conditions (C2) and (C3). Next, for (2.3) note that for  $\beta$  in a neighborhood of  $\beta_0$ , the functional derivative of  $H(\beta, g_{0\beta}, m_0, w_0, g'_{0\beta})$  in the direction  $[g - g_0, m - m_0, w - w_0, k - g'_0]$  for an arbitrary quadruple  $(g, m, w, k)$ , equals

$$\begin{aligned} & \Gamma(\beta, g_0, m_0, w_0, g'_0)[g - g_0, m - m_0, w - w_0, k - g'_0] \\ &= \lim_{\tau \rightarrow 0} \frac{1}{\tau} \left[ H(\beta, g_{0\beta} + \tau(g_\beta - g_{0\beta}), m_0 + \tau(m - m_0), w_0 + \tau(w - w_0), g'_{0\beta} \right. \\ & \quad \left. + \tau(k_\beta - g'_{0\beta})) - H(\beta, g_{0\beta}, m_0, w_0, g'_{0\beta}) \right] \\ &= E \left[ (w - w_0)(X) \left\{ \sigma^2(X, \beta_0, g_0) - \sigma^2(X, \beta, g_{0\beta}) \right\} \frac{\partial}{\partial \beta} \sigma^2(X, \beta, g_{0\beta}) \right] \\ & \quad - 2E \left[ w_0(X) \{Y - m_0(X)\} (m - m_0)(X) \frac{\partial}{\partial \beta} \sigma^2(X, \beta, g_{0\beta}) \right] \end{aligned}$$

$$\begin{aligned}
& - E \left[ w_0(X) \frac{\partial}{\partial z} \sigma^2(X, \beta, z) |_{z=g_{0\beta}(U)} (g_\beta - g_{0\beta})(U) \frac{\partial}{\partial \beta} \sigma^2(X, \beta, g_{0\beta}) \right] \\
& + E \left[ w_0(X) \left\{ \sigma^2(X, \beta_0, g_0) - \sigma^2(X, \beta, g_{0\beta}) \right\} \right. \\
& \quad \times \left\{ \frac{\partial}{\partial z_1} \sigma_\beta^2(X, \beta, z_1, g'_{0\beta}) |_{z_1=g_{0\beta}(U)} (g_\beta - g_{0\beta})(U) \right. \\
& \quad \left. \left. + \frac{\partial}{\partial z_2} \sigma_\beta^2(X, \beta, g_{0\beta}, z_2) |_{z_2=g'_{0\beta}(U)} (k_\beta - g'_{0\beta})(U) \right\} \right].
\end{aligned}$$

Note that the second term above equals zero. Hence, the first part of (2.3) follows easily from condition (C3) and (C4). For the second part, note that it follows from the proof of Theorem 2 in CLV that it suffices to show that

$$\begin{aligned}
& \|\Gamma(\beta, g_0, m_0, w_0, g'_0)[\hat{g} - g_0, \hat{m} - m_0, \hat{w} - w_0, \hat{g}' - g'_0] \\
& \quad - \Gamma(\beta_0, g_0, m_0, w_0, g'_0)[\hat{g} - g_0, \hat{m} - m_0, \hat{w} - w_0, \hat{g}' - g'_0]\| \\
& = o_P(1)\|\beta - \beta_0\| + O_P(n^{-1/2})
\end{aligned}$$

for all  $\beta$  with  $\|\beta - \beta_0\| = o(1)$ , and this follows easily from conditions (A1) and (A4). For (2.4), use condition (A1), (A2), (B1) and (B2) for the estimators  $\hat{g}$ ,  $\hat{g}'$  and  $\hat{m}$ . For  $\hat{w}$ , we showed above that  $\|\hat{w} - w_0\|_\infty = o_P(n^{-1/4})$ . Choosing  $\mathcal{W} = \{x \rightarrow \sigma^{-4}(x, \beta, g) : \beta \in B, g \in \mathcal{G}\}$ , it follows from assumption (A2) that  $P(\hat{w} \in \mathcal{W}) \rightarrow 1$  as  $n$  tends to  $\infty$ . Next, we consider (2.5). This condition can be checked by verifying the conditions of Theorem 3 in CLV. These follow from the Lipschitz continuity of the criterion function  $h$ , and from conditions (A2) and (B2). Moreover, using the differentiability of  $\sigma(x, \beta, z)$  with respect to  $\beta$  and  $z$ , it is easily seen that the covering number of  $\mathcal{W}$  is of the same order as that of  $B \times \mathcal{G}$ . Finally, for condition (2.6) note that when  $\beta = \beta_0$  all terms in the above calculation of the functional derivative cancel, except the third one. By inserting the expansion for  $\hat{g}_0 - g_0$  given in condition (A3) into this third term, and by using the notation  $T(u) = E\{w_0(X) \frac{\partial}{\partial z} \sigma^2(X, \beta_0, z) |_{z=g_0(u)} \frac{\partial}{\partial \beta} \sigma^2(X, \beta_0, g_0) | U_0 = u\}$ , we get:

$$\begin{aligned}
& \Gamma(\beta_0, g_0, m_0, w_0, g'_0)[\hat{g}_0 - g_0, \hat{m} - m_0, \hat{w} - w_0, \hat{g}'_0 - g'_0] \\
& = -(na_n)^{-1} \sum_{i=1}^n E \left[ w_0(X) \frac{\partial}{\partial z} \sigma^2(X, \beta_0, z) |_{z=g_0(U_0)} \frac{\partial}{\partial \beta} \sigma^2(X, \beta_0, g_0) \right. \\
& \quad \left. \times K_1 \left( \frac{U_0 - U_{0i}}{a_n} \right) \right] \eta(X_i, Y_i) + o_P(n^{-1/2}) \\
& = -(na_n)^{-1} \sum_{i=1}^n E \left[ T(U_0) K_1 \left( \frac{U_0 - U_{0i}}{a_n} \right) \right] \eta(X_i, Y_i) + o_P(n^{-1/2}) \\
& = -n^{-1} \sum_{i=1}^n T(U_{0i}) f_{U_0}(U_{0i}) \eta(X_i, Y_i) + o_P(n^{-1/2}),
\end{aligned}$$

where the latter equality follows from a Taylor expansion of order  $q$ .  $\square$



*Proof of Theorem 3.2.* Write

$$\begin{aligned}
& \hat{\sigma}^2(x) - \sigma^2(x, \beta_0, g_0) \\
&= \sigma^2(x, \beta_0, \hat{g}_{\hat{\beta}}) - \sigma^2(x, \beta_0, g_0) + O_P(n^{-1/2}) \\
&= \frac{\partial}{\partial z} \sigma^2(x, \beta_0, z)|_{z=g_0(u(x, \beta_0))} (\hat{g}_{\hat{\beta}} - g_0)(u(x, \beta_0)) \{1 + o_P(1)\} + O_P(n^{-1/2}) \\
&= \frac{\partial}{\partial z} \sigma^2(x, \beta_0, z)|_{z=g_0(u(x, \beta_0))} \left\{ (g_{0\hat{\beta}} - g_0) - (\hat{g}_0 - g_0) \right\} (u(x, \beta_0)) \{1 + o_P(1)\} \\
&\quad + O_P(n^{-1/2}),
\end{aligned}$$

where the latter equality follows from condition (A4). The result now follows from the representation for  $(\hat{g}_0 - g_0)(u)$  given in condition (A3) and from Theorem 3.1.  $\square$

*Proof of Theorem 3.3.* The proof is quite similar to that of Theorem 3.1. We focus here on the calculation of the derivative of  $H(\beta, g_{0\beta}, m_0, w_0)$  with respect to  $\beta$  and with respect to the nuisance functions, and on the verification of condition (2.5) in Chen, Linton and Van Keilegom (2003) (CLV). First, consider

$$\begin{aligned}
& H(\beta, g_{0\beta}, m_0, w_0, g'_{0\beta}) \\
&= E \left[ w_0(X) \left\{ 2P(|Y - m_0(X)| \geq \sigma(X, \beta, g_{0\beta}) | X) - 1 \right\} \frac{\partial}{\partial \beta} \sigma(X, \beta, g_{0\beta}) \right] \\
&= E \left[ w_0(X) \left\{ 1 - 2F(e_\beta(X, 1)|X) + 2F(e_\beta(X, -1)|X) \right\} \frac{\partial}{\partial \beta} \sigma(X, \beta, g_{0\beta}) \right],
\end{aligned}$$

where  $F(y|x) = P(Y \leq y | X = x)$  and  $e_\beta(x, y) = m_0(x) + \sigma(x, \beta, g_{0\beta})y$ , and hence

$$\begin{aligned}
& \frac{d}{d\beta^T} H(\beta, g_{0\beta}, m_0, w_0, g'_{0\beta}) \\
&= E \left[ w_0(X) \left\{ -2f(e_\beta(X, 1)|X) - 2f(e_\beta(X, -1)|X) \right\} \right. \\
&\quad \times \left. \frac{\partial}{\partial \beta} \sigma(X, \beta, g_{0\beta}) \frac{d}{d\beta^T} \sigma(X, \beta, g_{0\beta}) \right] \\
&\quad + E \left[ w_0(X) \left\{ 1 - 2F(e_\beta(X, 1)|X) + 2F(e_\beta(X, -1)|X) \right\} \frac{d}{d\beta^T} \frac{\partial}{\partial \beta} \sigma(X, \beta, g_{0\beta}) \right].
\end{aligned}$$

When  $\beta = \beta_0$ , the second term equals zero and we find the matrix  $\Lambda$  defined in the statement of the theorem. On the other hand,

$$\begin{aligned}
& \Gamma(\beta, g_0, m_0, w_0, g'_0)[g - g_0, m - m_0, w - w_0, k - g'_0] \\
&= E \left[ (w - w_0)(X) \left\{ 1 - 2F(e_\beta(X, 1)|X) + 2F(e_\beta(X, -1)|X) \right\} \frac{\partial}{\partial \beta} \sigma(X, \beta, g_{0\beta}) \right] \\
&\quad + E \left[ w_0(X) \left\{ -2f(e_\beta(X, 1)|X) + 2f(e_\beta(X, -1)|X) \right\} \right.
\end{aligned}$$

$$\begin{aligned}
& \times (m - m_0)(X) \frac{\partial}{\partial \beta} \sigma(X, \beta, g_{0\beta}) \Big] \\
& + E \left[ w_0(X) \left\{ -2f(e_\beta(X, 1)|X) - 2f(e_\beta(X, -1)|X) \right\} \frac{\partial}{\partial z} \sigma(X, \beta, z) \Big|_{z=g_{0\beta}(U)} \right. \\
& \quad \times (g_\beta - g_{0\beta})(U) \frac{\partial}{\partial \beta} \sigma(X, \beta, g_{0\beta}) \Big] \\
& + E \left[ w_0(X) \left\{ 1 - 2F(e_\beta(X, 1)|X) + 2F(e_\beta(X, -1)|X) \right\} \right. \\
& \quad \times \left\{ \frac{\partial}{\partial z_1} \sigma_\beta(X, \beta, z_1, g'_{0\beta}) \Big|_{z_1=g_{0\beta}(U)} (g_\beta - g_{0\beta})(U) \right. \\
& \quad \left. \left. + \frac{\partial}{\partial z_2} \sigma_\beta(X, \beta, g_{0\beta}, z_2) \Big|_{z_2=g'_{0\beta}(U)} (k_\beta - g'_{0\beta})(U) \right\} \right].
\end{aligned}$$

Note that the first and fourth term above equal zero when  $\beta = \beta_0$ . Finally, we consider the verification of condition (2.5) in CLV, for which we use Theorem 3 in that paper. To prove condition (3.2) in that theorem, we focus on the indicator in the criterion function  $h(x, y, \beta, g, m, w, k)$ , the other components (namely  $w(x)$  and  $\frac{\partial}{\partial \beta} \sigma(x, \beta, g)$ ) being much easier to deal with. Consider for any  $(\beta, g, m)$  (where  $\sup^*$  denotes the supremum over all  $\|\tilde{\beta} - \beta\| \leq \delta$ ,  $\|\tilde{g} - g\|_\infty \leq \delta$ ,  $\|\tilde{m} - m\|_\infty \leq \delta$ ),

$$\begin{aligned}
& E \left[ \sup^* \left| I(|Y - m(X)| - \sigma(X, \beta, g) \geq 0) - I(|Y - \tilde{m}(X)| - \sigma(X, \tilde{\beta}, \tilde{g}) \geq 0) \right|^2 \right] \\
& \leq \sup_x \left[ F(m(x) + \sigma(x, \beta, g) + \delta + \alpha|x) - F(m(x) + \sigma(x, \beta, g) - \delta - \alpha|x) \right] \\
& \quad + \sup_x \left[ F(m(x) - \sigma(x, \beta, g) + \delta + \alpha|x) - F(m(x) - \sigma(x, \beta, g) - \delta - \alpha|x) \right] \\
& \leq 4 \sup_{x,y} f(y|x)(\delta + \alpha),
\end{aligned}$$

where

$$\begin{aligned}
\alpha & := \sup^* \sup_x |\sigma(x, \beta, g) - \sigma(x, \tilde{\beta}, \tilde{g})| \leq \sup^* \{K_1 \|\beta - \tilde{\beta}\| + K_2 \|g - \tilde{g}\|_\infty\} \\
& \leq (K_1 + K_2)\delta,
\end{aligned}$$

for some  $0 < K_1, K_2 < \infty$ . Hence, condition (3.2) is satisfied for  $r = 2$  and  $s_j = 1/2$  (using the notation of CLV). Condition (3.3) in CLV follows easily from assumptions (A2), (B2) and (C3), where we choose  $\mathcal{W} = \{x \rightarrow \sigma^{-1}(x, \beta, g) : \beta \in B, g \in \mathcal{G}\}$ . The rest of the proof is similar to the one of Theorem 3.1 and is therefore omitted.  $\square$

## Acknowledgments

The first author acknowledges financial support from IAP research network P6/03 of the Belgian Government (Belgian Science Policy), and from the European Research Council under the European Community's Seventh Framework

Programme (FP7/2007-2013) / ERC Grant agreement No. 203650. The second author acknowledges support from the USA National Science Foundation grant DMS-0706842.

## References

- AKRITAS, M. G. AND VAN KEILEGOM, I. (2001). Nonparametric estimation of the residual distribution. *Scand. J. Statist.*, **28**, 549–568. [MR1858417](#)
- CARROLL, R. J. (2003). Variances are not always nuisance parameters. *Biometrics*, **59**, 211–220. [MR1987387](#)
- CHEN, X., LINTON, O. B. AND VAN KEILEGOM, I. (2003). Estimation of semi-parametric models when the criterion function is not smooth. *Econometrica*, **71**, 1591–1608. [MR2000259](#)
- CHIOU, J. M. AND MÜLLER, H. G. (2004). Quasi-likelihood regression with multiple indices and smooth link and variance functions. *Scand. J. Statist.*, **31**, 367–386. [MR2087831](#)
- DAVIDIAN, M. AND CARROLL, R. J. (1987). Variance function estimation. *J. Amer. Statist. Assoc.*, **82**, 1079–1091. [MR0922172](#)
- DAVIDIAN, M., CARROLL, R. AND SMITH, W. (1988). Variance functions and the minimum detectable concentration in assays. *Biometrika*, **75**, 549–556. [MR0967595](#)
- DELECROIX, M., HRISTACHE, M. AND PATILEA, V. (2006). On semiparametric M-estimation in single-index regression. *J. Statist. Plann. Infer.*, **136**, 730–769. [MR2181975](#)
- ENGLE, R. F. (1982). Autoregressive conditional heteroscedasticity and estimates of the variance of UK inflation. *Econometrica*, **50**, 987–1008. [MR0666121](#)
- GREENE, W. H. (2002) *Econometric Analysis*, (5th Edition). Prentice Hall, New Jersey.
- HÄRDLE, W., HALL, P. AND ICHIMURA, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.*, **21**, 157–178. [MR1212171](#)
- HÄRDLE, W., LIANG, H. AND GAO, J. (2000). *Partially Linear Models*. Physica-Verlag, Heidelberg. [MR1787637](#)
- HE, X. AND LIANG, H. (2000). Quantile regression estimates for a class of linear and partially linear errors-in-variables models. *Statist. Sinica*, **10**, 129–140. [MR1742104](#)
- HOROWITZ, J. L. AND LEE, S. (2005). Nonparametric estimation of an additive quantile regression model. *J. Amer. Statist. Assoc.*, **100**, 1238–1249. [MR2236438](#)
- KOENKER, R. AND ZHAO, Q. (1994). L-estimation for linear heteroscedastic models. *J. Nonpar. Statist.*, **3**, 223–235. [MR1291546](#)
- LEE, S. (2003). Efficient semiparametric estimation of a partially linear quantile regression model. *Econometric Theory*, **19**, 1–31. [MR1965840](#)
- LIANG, H., HÄRDLE, W. AND CARROLL, R. (1999). Estimation in a semiparametric partially linear errors-in-variables model. *Ann. Statist.*, **27**, 1519–1535. [MR1742498](#)

- MA, Y., CHIOU, J.-M. AND WANG, N. (2006). Efficient semiparametric estimator for heteroscedastic partially linear models. *Biometrika*, **93**, 75–84. [MR2277741](#)
- MÜLLER, H.-G. AND ZHAO, P.-L. (1995). On a semiparametric variance function model and a test for heteroscedasticity. *Ann. Statist.*, **23**, 946–967. [MR1345208](#)
- RUPPERT, D., WAND, M. P. AND CARROLL, R. J. (2003). *Semiparametric Regression*. Cambridge University Press. [MR1998720](#)
- RUPPERT, D., WAND, M. P., HOST, U. AND HÖSSJER, O. (1997). Local polynomial variance-function estimation. *Technometrics*, **39**, 262–273. [MR1462587](#)
- SCHICK, A. (1996). Weighted least squares estimates in partly linear regression models. *Statist. Probab. Letters*, **27**, 281–287. [MR1395580](#)
- VAN DER VAART, A. W. AND WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag, New York. [MR1385671](#)
- VAN KEILEGOM, I. AND CARROLL, R. J. (2007). Backfitting versus profiling in general criterion functions. *Statist. Sinica*, **17**, 797–816. [MR2398434](#)
- XIA, Y., TONG, H. AND LI, W. K. (2002). Single-index volatility models and estimation. *Statist. Sin.*, **12**, 785–799. [MR1929964](#)
- YATCHEW, A. (2003). *Semiparametric Regression for the Applied Econometrician*. Cambridge University Press.
- YU, K. AND JONES, M. C. (1998). Local linear quantile regression. *J. Amer. Statist. Assoc.*, **94**, 228–237. [MR1614628](#)
- ZHAO, Q. (2001). Asymptotically efficient median regression in the presence of heteroscedasticity of unknown form. *Econometric Theory*, **17**, 765–784. [MR1862375](#)